Topology of non-linear structure in the 2dF Galaxy Redshift Survey

J. Berian James,[1,2][★] Matthew Colless,[3] Geraint F. Lewis[2] and John A. Peacock[1]

[1]*Institute for Astronomy, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ*
[2]*School of Physics, the University of Sydney, Sydney, NSW 2006, Australia*
[3]*Anglo–Australian Observatory, PO Box 296, Epping, NSW 1710, Australia*

ABSTRACT

We study the evolution of non-linear structure as a function of scale in samples from the 2dF Galaxy Redshift Survey, constituting over 221 000 galaxies at a median redshift of $z = 0.11$. The two flux-limited galaxy samples, located near the Southern Galactic Pole and the Galactic equator, are smoothed with Gaussian filters of width ranging from 5 to $8\,h^{-1}$ Mpc to produce a continuous galaxy density field. The topological genus statistic is used to measure the relative abundance of overdense clusters to void regions at each scale; these results are compared with the predictions of the analytic theory, in the form of the genus statistic for (i) the linear regime case of a Gaussian random field and (ii) a first-order perturbative expansion of the weakly non-linear evolved field. The measurements demonstrate a statistically significant detection of an asymmetry in the genus statistic between regions corresponding to low- and high-density volumes of the Universe. We attribute the asymmetry to the non-linear effects of the gravitational evolution and biased galaxy formation, and demonstrate that these effects evolve as a function of scale. We find that neither analytic prescription satisfactorily reproduces the measurements, though the weakly non-linear theory yields substantially better results at some scales, and we discuss the potential explanations for this result.

Key words: galaxies: statistics – cosmology: observations – large-scale structure of Universe.

1 INTRODUCTION

The large-scale structure of the cosmological density field contains important information about the initial conditions and evolution of the Universe. This structure evolves solely under the influence of gravity, and its properties are a strong function of scale. Specifically, on the largest scales the morphology of the structure will reflect its primordial shape, expected to be that of nearly scale-invariant Gaussian random density perturbations (Bardeen, Steinhardt & Turner 1983; Peebles 1993; Peacock 1999), generated by the scalar field, or fields, responsible for cosmological inflation (Guth 1981; Albrecht & Steinhardt 1982; Liddle & Lyth 2000; Linde 2005). That this description accords with observation is strongly supported by galaxy redshift surveys (Geller & Huchra 1989; Shectman et al. 1996; Colless et al. 2001; York et al. 2000) and microwave background measurements (Bennett et al. 1994; Spergel et al. 2006), as well as the methods used to obtain information about the morphology of the large-scale structure (Gott et al. 1989; Melott et al. 1989; Gott et al. 1992).

Indeed, the tools for measuring these properties – the topology of the large-scale structure – are beginning to find wider application. It is more than 20 years since the publication of the seminal work

of Gott, Melott & Dickinson (1986), and the field has matured substantially. The initial question – whether the topology of the cosmological density field is consistent with inflation – has been answered in the affirmative through repeated analysis (Gott et al. 1989; Vogeley et al. 1994; Canavezes et al. 1998; Hikage et al. 2002; Hoyle et al. 2002; Hikage et al. 2003; Park et al. 2005; James, Lewis & Colless 2007), and studies of the departure from this as a result of the non-linear gravitational evolution, rather than as an initial condition, are an exciting new direction for observational studies (cf. Melott 1988 for pioneering work with simulations).

The classic topological statistic is the genus number of constant-density surfaces through the field, with the surfaces ordered so as to excise increasingly larger volumes. Each genus number is a measure of the connectedness of the structure at a particular density; the resulting locus, called the genus curve, is the genus number as a function of density relative to the median. The most important qualitative feature of this curve, for a Gaussian random field, is symmetry about this median density – the high- and low-density regions of the immediate post-inflation Universe are equivalent and interlocking, giving rise to a so-called *sponge topology*. With regard to the configuration of regions above and below median density, this is one of only three possibilities: it is the middle ground between the universe evoked by Press & Schechter (1974), a connected void populated with high-density clusters – a *meatball topology* – on the one hand and the single web of high-density structure punctured

[★]E-mail: jbj@roe.ac.uk

by voids – a *Swiss–cheese topology* – on the other hand. All three are idealizations: a spectrum runs smoothly from one extreme to the other, leading to the more common characterization in terms of meatball and Swiss–cheese *shift* away from the norm.

The genus curve has proven to be an especially valuable measure, as it can be predicted analytically for the case of the Gaussian random density field (Doroshkevich 1970; Adler 1981; Bardeen et al. 1986; Hamilton, Gott & Weinberg 1986). That the topology of structure will be scale-dependent should follow from the predictions of inflationary models together with observations of the local Universe, where isolated high-density structures predominate. This will produce a transition in the genus curve from symmetry at (sufficiently) large scales to a strong asymmetry in the non-linear regime of structure growth on small scale and at late times. Naively, such departures would be expected to arise at about the correlation length, where $\delta\rho/\rho \sim 1$ and modes of different sizes are beginning to interact. However, the form of the genus curve for a general field that does not obey Gaussian statistics has resisted analytical description, despite significant theoretical headway for specific cases of non-Gaussian fields (Hamilton 1988; Matsubara & Yokoyama 1996; Hikage et al. 2008) and second-order perturbation expansions of the Gaussian result (based on the methods of Kofman et al. 1994; Juszkiewicz et al. 1995) into the weakly non-linear regime (Matsubara 1994; Matsubara & Suto 1996; Seto et al. 1997). The latter case is especially salient, and Matsubara (2003) provides a key conclusion: that the effect of the weakly non-linear evolution on the genus curve is relatively modest.

It is against these expectations that the data must be tested. The weakly non-linear approximation will be observed to break down on sufficiently small scales, and the behaviour of the genus curves throughout and beyond the regime in which it is valid provides higher order information on the evolution of the density field than would be garnered with correlation functions or power spectra. Such a measurement requires a survey of a large volume with statistical properties that are well understood. The considerable sophistication of the statistical methods that have been brought to bear on galaxy redshift surveys since the turn of the century provides an opportunity to address this topic directly.

In this paper, we use the final data release of the 2dF Galaxy Redshift Survey (2dFGRS) to make measurements of the genus statistic on the cosmological density field on scales ranging from 5 to 8 h^{-1} Mpc. This builds upon the analysis of statistical properties of the galaxy distribution with this data set, including the power spectrum (Cole et al. 2005), void probability function (Croton et al. 2004b), the two-point and higher order correlation functions (Hawkins et al. 2003; Croton et al. 2004a) and the genus statistic measured in the linear regime as a test of the inflationary hypothesis (James et al. 2007). Section 2 describes how the density field is reconstructed from the galaxy survey and the restrictions this places on subsequent measurements. The genus measurement itself is the subject of Section 3, as are the physical meanings of the different qualitative results that may be obtained. Section 4 presents the results and method of error analysis for genus measurements of the 2dFGRS data, in which we test for the presence of a shift in the genus curve and discuss the implications of these results. Throughout, we use a cosmology in which $\Omega_{\rm m} = 0.27, \Omega_{\Lambda} = 0.73$ and $H_0 = 100\,h\,{\rm km\,s^{-1}\,Mpc^{-1}}$, with $h = 0.72$.

2 SAMPLES FROM THE 2DFGRS

At completion, the 2dFGRS (Colless et al. 2001) measured spectroscopic redshifts for 221 000 galaxies in the regions of the sky well separated to ensure statistical independence. With a limiting magnitude of 19.45 in the $b_{\rm J}$ band (Blair & Gilmore 1982), a median redshift of $\bar{z} = 0.11$ is obtained with close-to-uniform sampling across the sky. The survey comprises two regions located near the South Galactic Pole (SGP) and slightly to the north of the Galactic equator [North Galactic Pole (NGP)], along with a number of smaller random fields that will not be used for the analysis in this paper, giving a total survey area of over 2000 deg². The larger SGP region is a strip of approximately $90° \times 15°(21^{\rm h}40^{\rm m} < \alpha < 03^{\rm h}40^{\rm m}, -37°.5 < \delta < -22°.5)$ in which the redshifts of 115 492 galaxies are available for use in a statistically complete sample; the NGP strip is approximately $75° \times 10°(09^{\rm h}50^{\rm m} < \alpha < 14^{\rm h}50^{\rm m}, -7°.5 < \delta < +2°.5)$ and contains 79 696 galaxies.

2.1 Selection effects

The data reduction includes extinction and $k + e$ corrections, as well as masking of the survey area, to account for several distinct selection effects: (i) variation in magnitude limit between the two-degree fields that tile the full survey regions; (ii) variation between fields in the fraction of objects with high-quality spectroscopic redshifts and (iii) variations due to magnitude-dependent redshift completeness. These corrections are performed with the publicly available software of Peder Norberg and Shaun Cole, documented by Norberg et al. (2002). Information on the software, and the code itself, is provided on the 2dFGRS web site at http://www2.aao.gov.au/2dFGRS/.

Correcting for these effects addresses statistical incompleteness across right ascension and declination. When generating samples that maintain a constant number density with redshift, the effect of decrease in absolute magnitude limit with distance must be addressed. When the luminosity function of the population of observed galaxies is well understood, it is possible to estimate the fraction of galaxies – as a function of magnitude and redshift – that go unobserved as a result. The inverse of this number can be used to weight the galaxies that *are* seen, which serve as markers for the hidden population.

The combination of all statistical completeness effects is estimated with a single selection function, giving a weight to each galaxy. Because such a weighting scheme is useful for many different analyses of a single data set, the 2dFGRS selection function was computed by Norberg et al. (2002); they describe the estimators used to produce the weightings as well as other salient aspects of the $b_{\rm J}$-band luminosity function. Galaxy samples produced in this manner are *flux-limited* and retain all objects in the original catalogue at the expense of uncertainty over the exact values of the weighting and limitations imposed by the estimation.

An alternative to using a weighted galaxy sample is to remove the redshift-dependent effect by drawing a sample comprising just those objects that would be visible at the edge of the sample region, given the magnitude limit of the survey (e.g. Martínez & Saar 2002); such *volume-limited* samples are the more common way of making a measurement of the topological genus from a redshift survey (e.g. Gott et al. 1989; Canavezes et al. 1998; Hoyle, Vogeley & Gott 2002; Park et al. 2005; Gott et al. 2008); Vogeley et al. (1994) are notable counterexample. Flux-limited samples of the 2dFGRS were used by James et al. (2007), who demonstrated that (i) the choice of flux-limited catalogues reduced the noise in the measurement without introducing spurious effects and (ii) for the purposes of topological analysis, the selection function of the 2dFGRS is best restricted to the region $z < 0.2$.

2.2 Sample constraints

The choice of the maximum redshift to which a galaxy sample should extend depends on the properties of the underlying catalogue and the scale that is to be studied. This length-scale is imposed through the smoothing process, where it is identified with the characteristic width of the filter. The most restrictive requirement is that the smoothing length, λ, exceeds the average nearest-neighbour separation of galaxies in the sample volume. The latter quantity derives from the mean number density of the survey, $\bar{n}$; it is proportional to the cube root of the volume per object, $\bar{n}^{-1/3}$. For a Poisson point process, the mean nearest-neighbour separation is

$$\bar{\ell}_{nn} = \frac{\bar{n}^{1/3}}{\sqrt{\pi}} = \left(\frac{V}{\pi^{3/2}N}\right)^{1/3}, \tag{1}$$

where V is the volume of the survey and N is the number of galaxies. However, the cosmological density field is neither a continuous Poisson field nor the drawing of points from the background an unbiased process, so galaxies in redshift surveys should be assumed *not* to be a Poisson process, though the qualitative form of the relationship between $\bar{\ell}_{nn}$ and $\bar{n}$ should not change. A safe choice of smoothing length for a Gaussian filter is

$$\lambda > \bar{\ell}_{nn} \approx \left(\frac{V}{N}\right)^{1/3}; \tag{2}$$

this prescription has been used previously with success by Vogeley et al. (1994). The average number density is a function of the size of the survey; for a survey through a fixed window in right ascension and declination, $\bar{n}$ is a function of z_{max} only. In this way, the scale to be probed influences the choice of maximum redshift used when drawing samples from the 2dFGRS catalogue.

Another relationship between λ and z_{max} arises from considering the number of resolution elements in the smoothed sample. A high number of resolution elements is desirable in order to recover a statistically significant signal, though there is no critical threshold.

The number of resolution elements is

$$N_{res} = \frac{V_{survey}}{V_{kernel}} = \frac{V}{\pi^{3/2}\lambda^3} \Rightarrow \lambda = \left(\frac{V}{\pi^{3/2}N_{res}}\right)^{1/3}. \tag{3}$$

A value of $N_{res} = 10^3$ has been selected as indicative of a measurement of reasonable significance.

As has been stated, the choice of z_{max} is also restricted on account of the survey selection function; fig. 2 of James (2005) shows a break in the slope of the cumulative number density as a function of redshift at $z_{max} = 0.2$, leading to a natural choice for cutting off the maximum redshift at which an object should be included in this measurement. This defines a relation within the $z_{max}-\lambda$ plane, albeit one that is independent of λ. This plane, along with the constraints these three relations provide, is shown in Fig. 1.

Point samples are drawn from the NGP catalogue out to those values of z_{max} that will allow smoothing on scales of $\lambda = 5$ to $8\,h^{-1}$ Mpc in $1\,h^{-1}$ Mpc increments. The NGP catalogue has higher number density at low redshift, as it probes a smaller volume; the measurements in samples from the SGP catalogue will be correspondingly stronger and are also made above $5\,h^{-1}$ Mpc. Table 1 shows the salient parameters of each of the point samples drawn from the NGP and SGP slices.

2.3 Sample construction

The flux-limited galaxy samples from the NGP and SGP catalogues include all objects out to a maximum redshift given the constraints above. Initial input from the 2dFGRS catalogue takes the form of a list of redshift, right ascension and declination values, as well as weight for each galaxy corresponding to the inverse of the completeness. The redshift is converted to comoving distance (e.g. Hogg 1999) by numerically integrating

$$D_c = \frac{c}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_m(1+z)^3 + (1 - \Omega_m - \Omega_\Lambda)(1+z)^2 + \Omega_\Lambda}}; \tag{4}$$
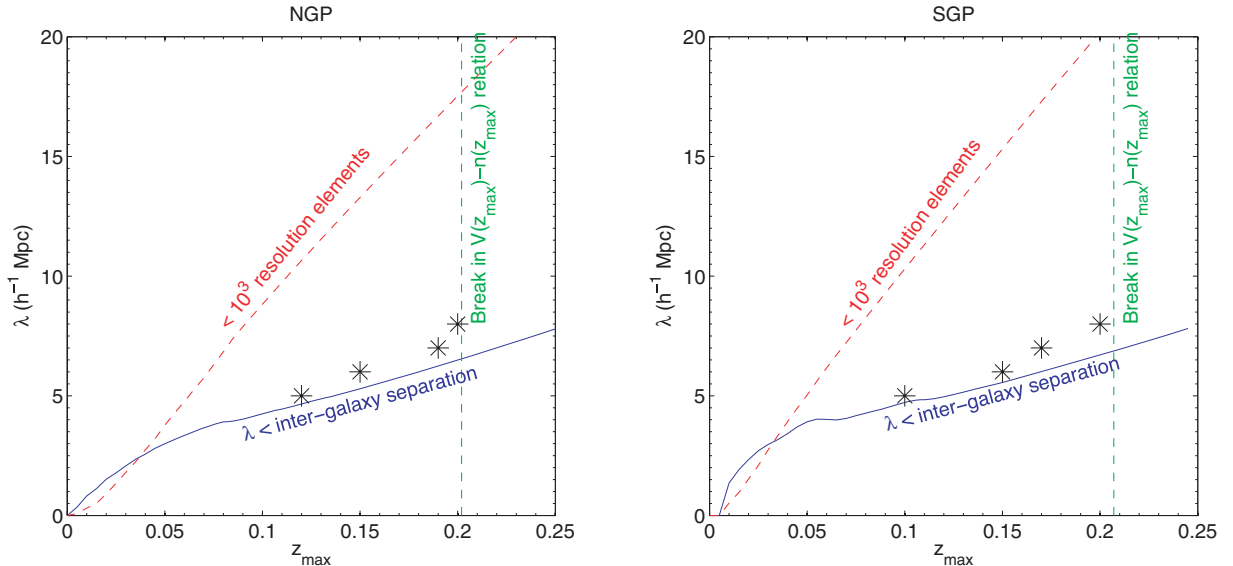

Figure 1. Constraint diagrams for the point samples drawn from the two 2dFGRS catalogues. Considerations of the selection function (dashed green), spatial number density (solid blue) and resolution (dashed red) restrict values of z_{max} and λ to the central triangular region. The constraint on the number of resolution elements is only indicative ($N_{res} = 10^3$); it is desirable to draw samples as close to the solid line as possible, so as to maximize the signal-to-noise ratio of the measurement while avoiding spurious signal from sparse points being treated as isolated clusters. The points indicate the choice of λ and z_{max} employed in the samples used in this work.

Table 1. Description of the point samples used in the genus measurement: for a given smoothing scale (λ, in h^{-1} Mpc), a value for the redshift limit (z_{max}) of the sample is chosen given the constraints of Fig. 1; this defines the volume (V, in cubic comoving megaparsecs), number of galaxies (N) and number density ($\bar{n}$) of the sample and the resolution of the array cells (in units of λ). The sum of galaxy weights (the 'power', a proxy for mass) is not conserved by the smoothing process, which pushes power outside the edges of the survey region. By normalizing the smoothed galaxy distribution to a smoothed constant field, no systematic effects are introduced by this process, though the loss of power ($p_{per\,cent}$) reduces the quality of the signal.

λ	z_{max}	V (/10^6)	N	$1/\bar{n}$	res	$p_{(per\,cent)}$
NGP						
5	0.12	6.652	48 892	136.0	0.20	11.44
6	0.15	12.98	63 431	204.6	0.25	10.31
7	0.19	25.74	73 897	348.3	0.29	9.33
8	0.20	29.86	75 268	396.7	0.25	10.18
SGP						
5	0.10	5.733	42 441	135.1	0.20	10.94
6	0.15	19.19	82 399	232.9	0.25	7.33
7	0.17	27.66	92 893	297.7	0.29	7.45
8	0.20	44.11	103 616	425.7	0.25	7.68

the distance and angular coördinates are converted into Cartesian coördinates, which are then rotated through the *y*-axis (corresponding to declination) so that the volume of the bounding box is minimized.

For the purpose of the smoothing and the genus measurement, the galaxies are represented within a three-dimensional array, binned using a nearest-grid-point algorithm, with resolution set to be well below the smoothing length; using bins of size $2\,h^{-1}$ Mpc is deemed adequate for the smoothing of $5\,h^{-1}$ Mpc or more. Figs 2 and 3 show the NGP and SGP slices embedded in this data structure. Cells outside the survey region must be flagged as such, rather than empty; they are given negative values.

However, by discretizing the spatial set of galaxy locations, those regions inside the slice removed during masking must also be flagged. If this is not done, these cells will act as false voids, leading to a Swiss–cheese signature in the genus measurement. Masking for the survey regions based on the composite survey completeness maps is applied by flagging regions of the α–δ plane within and around the NGP and SGP slices that are not part of the survey. The flagged regions are projected along the line of sight to define a pencil beam of flagged cells within the data array. The incursion of these cutouts around the edge of the slices (and the SGP slice in particular) is apparent in Figs 2 and 3.

The correctly masked data array is a three-dimensional structure with a flagged (negative) value in cells outside the survey bounds, as well as in masked regions within the bounds. The remaining cells have positive value and represent the weighted galaxy sample.

2.4 Sample smoothing

To retrieve an estimate of the cosmological density field, the galaxy sample is smoothed through a Gaussian window, though spherical top-hat (Hamilton et al. 1986) and wavelet techniques (Martínez et al. 2005) have also been considered for the purposes of the genus measurement. The width of the smoothing kernel acts to select a
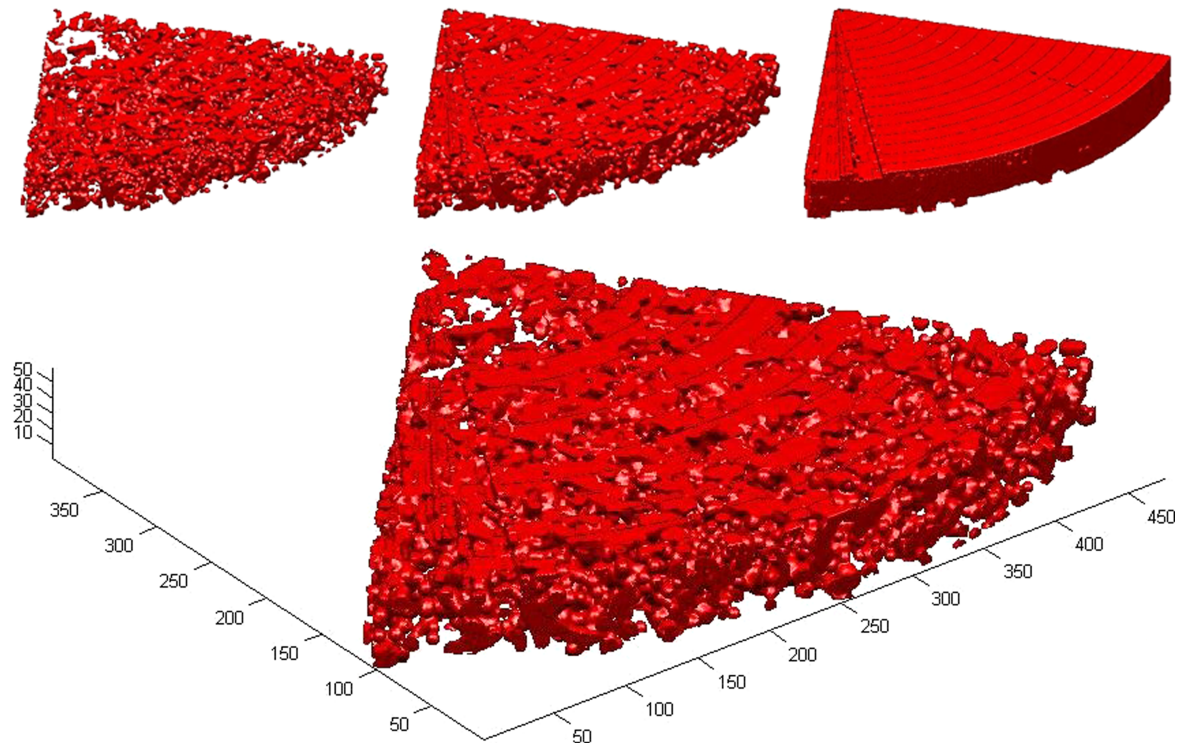

Figure 2. Surfaces of constant density through the cosmological density field of the NGP strip, extended out to a redshift of 0.2. The field is recreated by smoothing the statistically complete point catalogue with a Gaussian filter, in this case of width 4 Mpc, and subsequently embedding it within the data array, as demonstrated in the lower panel. The images in the upper panel show the changing topology of the surfaces as different volumes are excised: shown are surfaces containing (from top left) 33, 66 and 100 per cent of the volume within the sample. The single lower panel excises 50 per cent of the field – corresponding to the median density – showing the geometry of the NGP region within the $394 \times 475 \times 60$ array along with the effect of the survey mask. At the high-redshift end of the wedge, many isolated spherical surfaces are apparent, indicating that $z_{max} = 0.2$ is too large given the smoothing length (cf. Fig. 1).
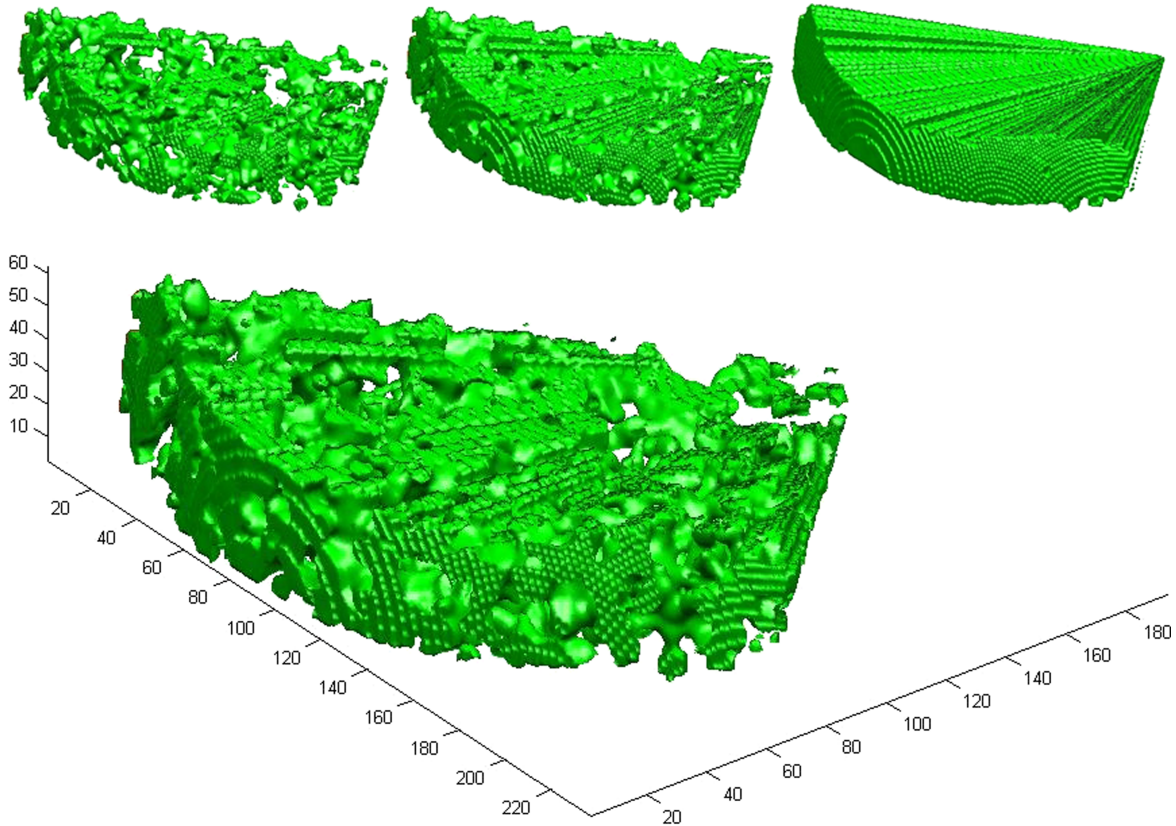
Figure 3. Isosurfaces of the SGP strip, extending to a redshift of 0.2, smoothing with a Gaussian of width $\lambda = 8\,h^{-1}$ Mpc. The upper panels again show the surfaces containing (from top left) 33, 66 and 100 per cent of the volume within the sample. Of note is the significantly smoother structures generated by the larger smoothing kernel; the topology is very close to that of a Gaussian random field.

scale, in Mpc, of the features of the field to be emphasized. As a result, by using kernels of a range of widths, the measurement is made across a range of scales. The appropriate range is governed by the statistical constraints of the sample, developed in Section 2.2, and by the aim of studying the non-linear regime. The transition to scales where the evolution of structure becomes non-linear, where modes of different scales become correlated, occurs at a characteristic length where the rms amplitude of density perturbations is of the same order as the background, i.e. where $\delta\rho/\rho \sim 1$. Measurements of the variance of this fractional rms amplitude between spheres of radius $8\,h^{-1}$ Mpc, σ_8, have converged about 0.9 (Tytler et al. 2004), though disagreement persists (*e.g.* Spergel et al. 2006). Here, the same value as other clustering analyses with the 2dFGRS data set is used: $\sigma_8 = 0.89$. To study the weakly non-linear regime, therefore, it is necessary to probe around $8\,h^{-1}$ Mpc; based on the sampling constraints outlined in Section 2.2, kernel widths of 5, 6, 7 and $8\,h^{-1}$ Mpc have been chosen for both regions. These are related to the spherical top-hat filter radius, which corresponds more precisely to the notion of physical distance, by $\lambda_G = \lambda_T/\sqrt{5}$; this result follows from the conservation of the filter volumes, and the distinction should be kept in mind when interpreting the results.

The smoothing is carried out via Fourier-space convolution[1] of the sample with a Gaussian kernel,

$$G(\boldsymbol{x}) = \exp\left(-\frac{\boldsymbol{x}' \cdot \boldsymbol{x}'}{2\lambda^2}\right); \tag{5}$$

[1] We use the FFTW libraries (Frigo & Johnson 2005,, v3.1.2 from http://www.fftw.org) for the discrete Fourier transform.

as the survey regions do not completely, or even nearly, fill the data volume – the slice geometry occupies approximately one-third of a tightly fitting rectangular prism – there will be power loss through the outer surfaces of the slice from galaxies smoothed partially into the flagged non-survey region. This signal is not recoverable and creates a relative drop in power near the slice boundaries that will, if left untreated, produce a volumetric overabundance of low-density region, leading to a Swiss–cheese signature. To avoid this, the density must be artificially raised in a manner that does not disturb the topological measurement.

Melott & Dominik (1993) have demonstrated a method that corrects for this systematic near the boundaries: create a three-dimensional array mirroring the geometry of the sample, but filled with a constant-density field (say of value $\rho = 1$) rather than sparsely separated galaxies. Smoothing this constant data array mimics the power loss through the surfaces of the survey boundary; dividing the smoothed 2dFGRS data array cell-wise by the smoothed constant data array will remove this effect. Furthermore, they demonstrate that the only systematic effect induced by this process is a slight increase in the amplitude of the measured genus curve, a correction proportional to the ratio of surface area to volume that is negligible in this context; moreover, there is no change in the symmetry properties of the curve that are studied in this work. A more important penalty is the decrease in signal – when λ is even a moderate percentage of the survey size, the fraction of signal lost can be quite large; these values are given in the last column of Table 1.

3 TOPOLOGICAL ANALYSIS

The morphology of the galaxy distribution has long been known to convey information from all orders of the correlation function, albeit in a way that is yet to be completely elucidated. Unlike geometric measures of structure, morphological statistics concentrate on properties such as 'connectedness' and global curvature rather than on orientation or size. The most famous of these, the Minkowski functionals (Mecke, Buchert & Wagner 1994; Kerscher, Schmalzing & Buchert 1996; Schmalzing & Buchert 1997), give a concise description of the density field up to topological equivalence; for two-dimensional surfaces embedded in three-dimensional space, they correspond to the enclosed volume, area, integrated mean and Gaussian curvatures of the surface.

The latter is proportional to the number of holes through the surface, called the genus number; informally

$$g = \text{number of holes} - \text{number of isolated regions} + 1. \quad (6)$$

When a surface is constructed through a three-dimensional field on the basis of a physical quantity, the genus number measures how connected (when $g > 0$) or disjoint ($g < 0$) regions demarcated by this quantity are. The quantity of interest to cosmologists is mass density; for instance, if the surface is drawn at the critical density for spherical collapse, the genus number will measure how interlinked such regions are relative to the background.

The numerical calculation of the genus number is done indirectly by computing the integrated curvature for surfaces through an array containing the smoothed density field. The algorithm of Weinberg, Gott & Melott (1987) marks regions as either above or below a critical density and defines the surface as the boundary between them. The relationship between the genus number and the integrated Gaussian curvature (on a smooth manifold) is the Gauss–Bonnet theorem:

$$\int_S K \, dA + \int_{\partial S} k \, ds = 4\pi(1 - g), \quad (7)$$

where K is the Gaussian curvature and k is the curvature on the boundary of the surface; the second integral vanishes whenever the surface is closed. Because the computation of the genus number takes place in a polygonal data structure, the surface is composed of the flat sides of array cells along with the vertices and edges that join them; it is the result of an earlier theorem of Descartes (Adam & Tannery 1996; Sasaki 2003) on the angle defect[2] of polyhedra that is relevant, *viz.*

$$4\pi(1 - g) = \sum \left(2\pi - \sum_i A_i \right), \quad (8)$$

where A_i is the angle made by each face of the cell that meets a vertex and the outer sum is evaluated over every vertex in the array. This is the quantity that must be evaluated for each cell vertex of the array containing the smoothed density field; the angle deficit matrices tabulated by Weinberg et al. (1987) are now used to compute the genus number of the surface.

3.1 Genus curve of embedded surfaces

The measurement of genus number is made not just on one surface, but on a sequence of surfaces spanning the range of densities con-

[2] Cf. the more modern sounding 'angle deficit'; while the description of this measurement as an application of the Gauss–Bonnet theorem – which applies specifically to smooth manifolds – is hardly inappropriate, the more general index theorem of Atiyah & Singer (1963) is also worthy of mention.

tained within the field. The characteristic density that defines each surface is chosen to excise a fraction of the volume of the field, parametrized by the parameter ν,

$$v_f(\nu) = \frac{1}{\sqrt{2\pi}} \int_\nu^\infty e^{-t^2/2} \, dt = \frac{1}{2} \text{Erf}_c \left(\frac{\nu}{\sqrt{2}} \right), \quad (9)$$

where Erf_c is the conjugate error function. In this way, $\nu = 0$ corresponds to splitting the field into the higher and lower mass regions such that each occupy 50 per cent of the volume; roughly the highest density 15 per cent of the volume lies above the surface at $\nu = \pm 1$, 2 per cent for $\nu = 2$ and so on as for the usual Gaussian integral.

Defining the value of the density that splits the field into the proper volume fraction is done by reshaping the three-dimensional density field into an ordered one-dimensional array so that, as each element of the array represents a standard unit of volume, moving p per cent of the way along the array will give the critical density corresponding to enclosing $(100 - p)$ per cent of the volume. That is ρ_c is the nth element of the sorted arrays for

$$n = \text{floor}[(1 - v_f)N], \quad (10)$$

where N is the length of the ordered array.

Excising the high- and low-density regions with this value and measuring the genus number on the intervening surface gives a curve as a function of the volume parameter ν, i.e. $\nu \rightarrow v_f \rightarrow \rho_c \rightarrow g$. The canonical shape of this curve for a Gaussian random field (formulated below in equation 11) is shown in Fig. 4: when ν is negative, only small disjoint regions of the field are excised, so the genus number at this value is negative; as ν increases, more disjoint regions appear and the genus number becomes even lower; eventually these regions start to join and the genus number of the surfaces increase, crossing at $\nu = -1$; the curve peaks at $\nu = 0$ (median density), where the field is maximally connected; the
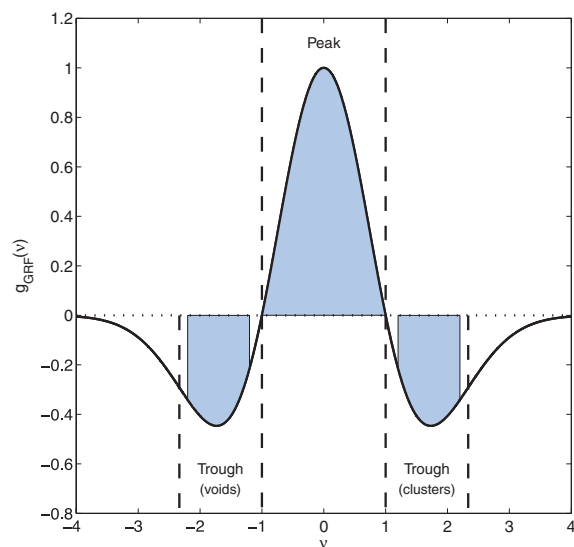

Figure 4. The archetypal genus of the Gaussian random field (solid line) normalized to the peak value at $\nu = 0$ and (demarcated by dashed lines) the regions of distinct topological character corresponding to the presence of isolated voids and clusters, for the troughs and for the connected surface at median density. Interpretations of genus curve measurements focus on differences between the observed curve and that of an idealized model, such as the Gaussian case. Departures from the model are quantified using the ratio between model and measurement of the area under three regions of the curve (shaded) (this is discussed in Section 3.3).

curve then recedes symmetrically – for the case of a Gaussian field, high- and low-density regions are connected in the same way.

So, there are three domains throughout the curve: (i) $|\nu| < 1$, the 'central peak', where the genus curve is positive, corresponding to those fractions of the volume that are more connected than isolated; (ii) $1 \le |\nu| \lesssim 2.5$, the 'troughs', where the genus curve is negative and (iii) $|\nu| \gtrsim 2.5$, the 'wings', where the value of the curve, though still negative, is trending asymptotically to zero. While the crossing point from positive to negative values of genus number serves to separate the central peak from the troughs, the point of demarcation between the troughs and the wings is less clear and somewhat more arbitrary; the points of inflection at $\nu = \pm\sqrt{3 + \sqrt{6}} \approx \pm 2.33$ are the obvious candidates for a rigourous division, but this is perhaps unnecessary. While such a division provides little direct insight, it has provided the implicit basis for statistical analysis of the genus curve since at least Park, Gott & da Costa (1992); the wings, as a region of especially low signal-to-noise ratio, are often excluded from measurement, while the interplay between the central peak and the troughs guides physical interpretation.

3.2 Modelling of the genus curve

To properly determine the physical significance of genus measurement made on data from redshift surveys, models of the genus curve itself are required. So far, theoretical success has been restricted to the special case of the Gaussian random field – believed to accurately represent the statistical properties of the density field immediately following the end of cosmological inflation – and small departures from this case, such as would be expected under smooth gravitational evolution.

Consequently, these two descriptions are appropriate in different regimes: on large scales, at least $\lambda = 10\,h^{-1}$ Mpc, the genus curve equation for a Gaussian field tests whether the observed density field is faithfully captured by models of the early Universe. The perturbative model applies where small departures from the Gaussian case are expected, i.e. on scales below, but not much below, the divide between the linear and the non-linear regimes. It is these scales that are probed by the current work.

3.2.1 Linear regime: Gaussian random field

Much of the early success of the genus curve as a diagnostic tool for diverging models of structure formation can be attributed to the construction of an analytic function for the special case of a Gaussian random field. As calculated by a number of independent researchers (Doroshkevich 1970; Adler 1981; Bardeen et al. 1986; Hamilton et al. 1986), the formula is

$$g_{\text{GRF}}(\nu) = A \exp(-\nu^2/2)(1 - \nu^2); \tag{11}$$

where the amplitude of the curve is determined by the power spectrum of the field (Hamilton et al. 1986):

$$A = \frac{1}{4\pi^2} \left(\frac{\langle k^2 \rangle}{3} \right)^{3/2}, \tag{12}$$

where

$$\langle k^2 \rangle = \frac{\int k^2 P(k) W^2(k\lambda) d^3 k}{\int P(k) W^2(k\lambda) d^3 k}. \tag{13}$$

When the window function W is a Gaussian of width λ – such as equation (5) – and the power spectrum $P(k)$ a power law with

spectral index n, the normalization factor is

$$A = \frac{1}{4\pi^2 \lambda^3} \left(\frac{3 + n}{3} \right)^{3/2}. \tag{14}$$

The significance of this result is that the curve is necessarily symmetric for general choices of power spectrum; the result is applicable even for anisotropic fields; Matsubara & Suto (1996) have shown that the impact of redshift-space distortions in the linear and weakly non-linear regimes is small. The high- and low-density regions of the field are treated equally with respect to the volumes they excise and to the connectedness of such regions. While the term 'Gaussian random field' refers to the case where the Fourier amplitudes of the field are Rayleigh distributed and the phase distribution is uniform in the interval $[0, 2\pi)$, it is the latter condition that is most salient when probing large volumes of the Universe; even fields with power distributed in a deterministic fashion [e.g. $P(k) \propto k^n$] are often referred to as 'Gaussian' when they satisfy the random phase condition; this amounts to an implicit invocation of the central limit theorem.

3.2.2 Weakly non-linear regime: the perturbation theory

Correlations between phases such as those induced by non-linear gravitational evolution lead to the disruption of symmetry about the median density contour of the genus curve. The nature of this evolution has been described in a series of papers (Matsubara 1994; Matsubara & Suto 1996) using the cosmological perturbation theory: in a Gaussian field δ, all connected moments higher than $n = 2$ vanish; in a non-Gaussian field satisfying a technical requirement in which higher moments are of the order of a specific power of the rms fluctuations of the field σ, the genus curve is given by

$$g_{\text{WNL}}(\nu) \approx -A \exp(-\nu^2/2)$$
$$\times \{H_2(\nu) + \sigma \left[P_3 H_3(\nu) + P_1 H_1(\nu) \right]\}, \tag{15}$$

where $H_n(\nu) \equiv (-1)^n \exp(\nu^2/2)(\mathrm{d}/\mathrm{d}\nu)^n \exp(-\nu^2/2)$ are the Hermite polynomials, their coefficients P_n are determined from the skewness parameters $S^{(a)}$ of the smoothed field (defined as *per* equation 2.8 of Matsubara & Suto 1996); the approximation to the genus formula in equation (15) arises only from discarding terms of the order of σ^2 and above; and the normalization A is unchanged from the linear regime. It is this equation that can be used to predict the genus curve on scales below the linear regime, and it has been used in the analysis of the results.

The genus formula for the Gaussian random field is expressible in this formalism as well, *viz.*

$$g_{\text{GRF}} = -A \exp\left(-\nu^2/2\right) H_2(\nu), \tag{16}$$

where now, with only an even Hermite polynomial present, the genus curve is symmetric about $\nu = 0$ – notably, this will not be the case in the weakly non-linear regime. The value of the skewness parameters, as well as the rms fluctuation in amplitude of the field, changes as a function of scale. The Hermite coefficients are simply given by

$$P_3 = S^{(1)} - S^{(0)} \quad \text{and} \quad P_1 = S^{(2)} - S^{(0)}; \tag{17}$$

over the range of scales probed in this work, the values of the skewness parameters – and hence those of the Hermite coefficients – can be fit by a cubic polynomial. Table 2 shows the polynomial coefficients of the parameters, while the value of σ can be left as a free parameter.

Table 2. All three skewness parameters are estimated by the function $S(\lambda) = \sum_{n=0}^{3} s_n \lambda^n$; this table shows the coefficients s_n for each skewness parameter, as well as the Hermite coefficients of equation (15).

	$s_3(\times 10^{-6})$	$s_2(\times 10^{-3})$	$s_1(\times 10^{-3})$	s_0
$S^{(0)}$	-98.96	4.625	-88.42	4.007
$S^{(1)}$	-101.6	4.750	-92.63	4.104
$S^{(2)}$	-20.83	0.9063	-8.792	3.681
P_3	-2.604	0.1250	-4.208	0.09700
P_1	78.13	-3.719	79.63	-0.3260

Figure 5. Fractional difference between the bootstrap and the Poisson estimates of errors for the genus of a simulated Gaussian random field subject to the same geometric constraints as the data in this work.

3.3 Physical metastatistics

Both of the models in Section 3.2 can be used as analytic fits to genus measurements from the 2dFGRS data. As the range of scales considered in this work pass well into the non-linear regime, it is to be expected that the genus measurement will deviate substantially from equation (11).

A measure of such departures was provided by Park et al. (1992), who define the *genus shift* parameter,

$$\Delta\nu(\lambda) = \frac{\int_{-1}^{1} \nu g(\nu; \lambda) d\nu}{\int_{-1}^{1} g_{\text{model}}(\nu; \lambda) d\nu}; \tag{18}$$

when the model accurately reproduces the genus measurement, $\Delta\nu = 0$, but when a meatball or Swiss–cheese shift is recorded, $\Delta\nu$ becomes greater or less than unity, respectively. The scale dependence of the parameter is significant because the non-linear effects induced by gravitational evolution cause a modest meatball shift in the genus curve (Park & Gott 1991).

To determine which regions of the genus curve account for such deviations – and thus provide motivation for interpretation of the results in terms of voids and clusters – further metastatistics have been defined by Park, Gott & Choi (2001) and Park, Kim & Gott (2005),

$$A_v(\lambda) = \frac{\int_{-2.2}^{-1.2} g(\nu; \lambda) d\nu}{\int_{-2.2}^{-1.2} g_{\text{model}}(\nu; \lambda) d\nu} \text{ and} \tag{19}$$

$$A_c(\lambda) = \frac{\int_{1.2}^{2.2} g(\nu; \lambda) d\nu}{\int_{1.2}^{2.2} g_{\text{model}}(\nu; \lambda) d\nu} \tag{20}$$

that probe the deviation from the model in the regions about the troughs of the genus curve below and above $\nu = 0$, respectively, in contrast to $\Delta\nu$, which is localized about the central peak.

4 RESULTS

We measure the genus statistic at 100 points in the range $-4 \leq \nu \leq 4$ for the NGP and SGP slices smoothed on scale ranging from 5 to $8\,h^{-1}$ Mpc. The errors on the measurements are estimated using the technique of bootstrap resampling (Efron 1982; Barrow, Bhavsar & Sonoda 1984), in which we create distinct samples with the same variance properties as the data by redrawing – with replacement – individual galaxies from the input 2dFGRS catalogue up to the full number of objects; the use of jackknife resampling has not yet been properly explored in the context of the genus measurement. After creating 100 such resampled catalogues, each one is subjected to the method described in Section 2.

To test the robustness of the bootstrap resampling, we simulate a Gaussian random density field, with the same geometry as the survey data, from which 100 Poisson samplings are drawn. These
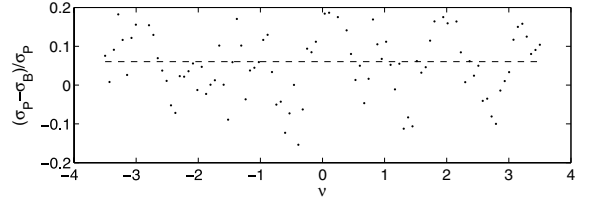
are smoothed as *per* the data and the genus of the resulting distributions measured. In Fig. 5, the variance of these is compared to the variance of the genus measurements from the same number of bootstraps of a single Poisson sample. It is found that, although the bootstrap resamples display on average a meatball shift, the variance between them is unbiased – on average, the error bars estimated from the bootstrap are 6 per cent smaller than those from the Poisson samples, but do not exhibit any large-scale ν-dependent trends. Though there is a considerable ($\sim$10 per cent) tendency for the bootstrap to both over- and underestimate the error bars in the regions to which Δ_ν, A_ν and A_c are sensitive, this scatter is not systematic; we conclude that the use of bootstrap resampling along with this smoothing method will not create anomalous values for the metastatistics by virtue of distorting the error bars on the genus measurement.

From the resulting set of mock genus measurements, the covariance matrix of the genus curve is computed for both slices and at each smoothing scale. Recent work (Hartlap, Simon & Schneider 2007) has demonstrated that the bootstrap is a biased estimator of the inverse covariance matrix: the location of the peak in the likelihood function is unchanged – so the best-fitting model is correct irrespective – but the slope of the likelihood function in the region about the peak will be overestimated, making the fit appear better than it is; a scaling of the inverse correlation matrix, in accordance with equation (17) of Hartlap et al. (2007), produces an unbiased estimator.

This provides a way to perform a χ^2 analysis of fits to the measurement. For the two models discussed in Section 3.2 (equations 11 and 15), we fit the genus measurement by minimizing the resulting dispersion

$$\chi^2 = \frac{1}{n - p} [g(\nu) - g_{\text{model}}(\nu)]^{\text{T}} \mathbf{C}^{-1} [g(\nu) - g_{\text{model}}(\nu)], \tag{21}$$

where $\mathbf{C}^{-1}$ is the rescaled covariance matrix, $n = 100$ is the number of data points and p is the number of model parameters and is either one or two.

In this section, we compare the 2dFGRS observations with the models and examine trends with scale, as well as differences arising from the choice of model and metastatistics; these results are related back to the distributions of voids and clusters.

4.1 2dFGRS topology below the linear regime

The genus measurements for the NGP and SGP slices are shown in Fig. 6, along with the best-fitting Gaussian random field and weakly perturbed model genus curves. When calculating statistics such as $\Delta\nu$, the model fitting is done by minimizing the scatter in the region $-3 \leq \nu \leq 3$, rather than the traditional $-1 \leq \nu \leq 1$, in order to use information from all parts of the genus curve. The values of the metastatistics described in Section 3.3 for the NGP and SGP slices,
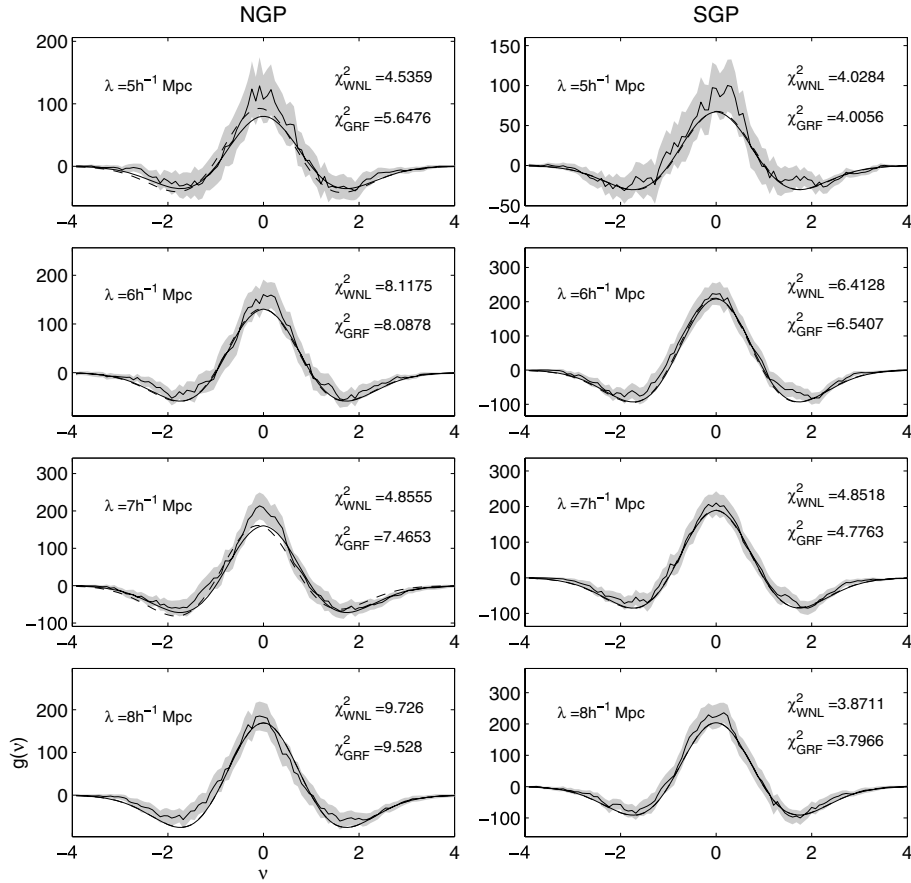
Figure 6. Genus curves for samples drawn from the NGP (left-hand panels) and SGP slices of the 2dFGRS over scales from 5 to $8\,h^{-1}$ Mpc, with 3σ uncertainty region estimated from the variance between bootstrap resamples of the data sets. The two fits to the measurements are made with the models of equations (11) (smooth solid line), leaving the amplitude of the curve as a free parameter, and (15) (dashed line), with both the amplitude and density rms σ varying; in several cases the two curves are nearly identical. The goodness-of-fit values quoted are the reduced χ^2, accounting for the additional parameter in our fit of the weakly non-linear model.

Table 3. Asymmetry metastatistics for the 2dFGRS samples measured relative to the best-fitting genus curve for a Gaussian random field as *per* equations (18), (19) and (20). 1σ confidence limits are the standard deviation of the set of mock asymmetry statistics generated using the resampled genus measurements, relative to the same model curve.

λ	$\Delta\nu$	A_c	A_v
NGP			
5	0.0063 ± 0.0277	0.9957 ± 0.1113	0.8534 ± 0.0840
6	0.0158 ± 0.0167	0.8056 ± 0.0464	0.7400 ± 0.0388
7	-0.0399 ± 0.0166	0.8997 ± 0.0508	0.7378 ± 0.0433
8	-0.0598 ± 0.0150	0.7693 ± 0.0438	0.6348 ± 0.0336
SGP			
5	0.0070 ± 0.0230	0.6741 ± 0.0576	0.8143 ± 0.0670
6	-0.0249 ± 0.0129	0.7811 ± 0.0337	0.7569 ± 0.0384
7	-0.0060 ± 0.0143	0.8850 ± 0.0360	0.8054 ± 0.0361
8	-0.0295 ± 0.0148	0.9926 ± 0.0412	0.8040 ± 0.0367

taken relative to the best-fitting genus curve of a Gaussian random field (equation 11), are given in Table 3.

These measurements show that for both the NGP and SGP samples $\Delta\nu$ deviates below zero, indicating that the samples are cluster-dominated at some scales (a meatball-shift topology); we can test

the significance of this deviation using the statistic

$$\chi^2_{\Delta\nu} = \frac{1}{4} \sum_{\lambda=5/h}^{8/h} \left[\frac{\Delta\nu(\lambda) - \Delta\nu_{GRF}}{\sigma_{\Delta\nu}(\lambda)} \right]^2 \tag{22}$$

for each of the samples, which measures the deviation from the Gaussian value $\Delta\nu_{GRF} = 0$, and which is tabulated within Table 4. This shows that, assuming that the uncertainties are normally distributed, the topology of the SGP region is unlikely to be consistent with a Gaussian random field and that the NGP is definitely not to a high level of significance. This is in agreement with the results of simulations presented by Park et al. (2005) for the case of biased peaks of the Λ cold dark matter density field, and inconsistent with other cases including unbiased dark matter and haloes.

The measurements also indicate that both A_c and A_v are below unity, showing a lower multiplicity of clusters and voids relative to the case of the Gaussian random field. Table 4 again quantifies the departure from the Gaussian case, using the analogous statistics to equation (22), where the deviation is now from unity rather than zero. Park et al. (2005) have demonstrated how the value of A_v might descend below unity through (linearly) biased galaxy formation. It is unclear what general mechanism can drive A_c down as well, though a physical argument is that when the average nearest-neighbour separation of the sites of galaxy formation is shorter than the Gaussian case, at a fixed smoothing scale there will be less

Table 4. One-sample (reduced) χ^2 tests on the statistics derived from the genus curve measurements, testing in the first place the case that the density field at these scales is a Gaussian random field (the parameters have their Gaussian values $\{\Delta\nu, A_c, A_v\} = \{0, 1, 1\}$), and in the second place that the topology of the density field is not evolving with scale (the parameters have their constant best-fitting values).

	Against the hypothesis of a Gaussian random field			Against the hypothesis of no evolution		
	$\chi^2_{\Delta\nu}$	$\chi^2_{A_c}$	$\chi^2_{A_v}$	$\chi^2_{\Delta\nu}$	$\chi^2_{A_c}$	$\chi^2_{A_v}$
NGP	5.66	12.3	50.74	3.35	1.57	2.21
SGP	2.00	21.13	26.33	0.69	6.60	0.29

isolated high-density peaks in the field. This is accompanied by greater connectivity of void regions, resulting in a decline in A_v; whether this need always be the case is not clear.

4.2 Scale-dependent evolution

There are trends in each of $\Delta\nu, A_c$ and A_v as a function of scale shown in Fig. 7. The trend in the value of $\Delta\nu$ over the scales examined from negative to zero (within the uncertainty) is in partial agreement with the measurements on simulations by Park et al. (2005) who detect a trend from positive to negative as λ increases, with a crossing point at $6\,h^{-1}$ Mpc. We observe no evidence of such a crossing in the SGP slice, though Matsubara & Suto (1996) show that the effect of velocity-space distortions on the small-scale genus curve suppresses the value of $\Delta\nu$ in the weakly non-linear regime. The value of $\Delta\nu$ in the NGP slice more accurately matches the predictions of both theory and simulation, moving from below zero to zero in the range between 7 and $6\,h^{-1}$ Mpc.

Table 4 also gives the reduced χ^2 for the best-fitting non-evolving values of the parameters. Though the two samples do not agree on trends in A_c and A_v, it is clear that the topology is evolving with scale to a high statistical significance. The most remarkable trend is in the SGP slice, which displays a cluster multiplicity consistent with a Gaussian random field on the largest scales and decreasing smoothly to well below Gaussian levels at smaller scales. The source of this trend is clearly discernible in Fig. 6 as a relative elevation in the genus curves over the region in which the cluster multiplicity integral is calculated. Moreover, this elevation broadens and shifts to larger values of ν on smaller scales, indicating an increasing degree of connection of high-density clusters, which (i) occupy an increasingly extreme fraction of the sample volume and (ii) extend across a widening range of densities. To complement this, the void multiplicity is essentially stable at subGaussian field levels.

The NGP slice also displays evolution, though the trends are less clear: while the void multiplicity appears to increase towards smaller scales, this could be a result of suppression in the number
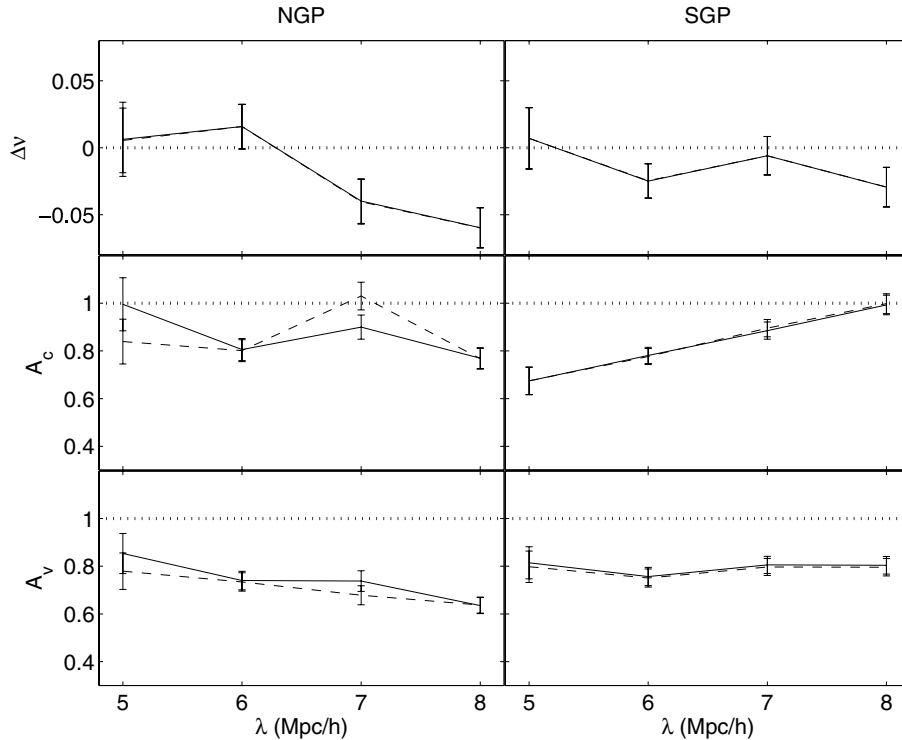

Figure 7. Genus curve metastatistics $\Delta\nu$, A_c and A_v as a function of scale for the 2dFGRS NGP and SGP survey slices, with 1σ error bars determined by the variance between the same quantities calculated on each of the bootstrap resamples. Each quantity is calculated by taking the ratio between the data and a model the area under different regions of the genus curve (NB. However, a factor of ν in the integrand of the numerator in equation 18). Results are shown for the calculation relative to the genus curve for a Gaussian random field (solid line) and for its weakly non-linear perturbation (dashed), with the null hypothesis Gaussian random field parameter values (dotted) shown for reference.

of isolated voids at *larger* scales due to large, connected overdense structures within the NGP field. It is possible to test for the two samples describing the same evolution in each parameter using the statistic

$$\chi^2_{\text{cross}} = \frac{1}{4} \sum_{\lambda=5/h}^{8/h} \left[\frac{\Delta\nu_{\text{NGP}}(\lambda) - \Delta\nu_{\text{SGP}}(\lambda)}{\sqrt{\sigma_{\text{NGP}}(\lambda)^2 + \sigma_{\text{SGP}}(\lambda)^2}} \right]^2, \tag{23}$$

where each σ is the error bar on the $\Delta\nu$ parameter and where analogous statistics can be defined for A_c and A_v. The NGP and SGP samples do not show consistent trends and offsets in the parameter values, with the cross-statistic yielding 2.05 ($\Delta\nu$), 5.16 (A_c) and 3.30 (A_v). On the scales probed in this work, it appears that the topology of structure differs modestly between the two regions at median density and more strongly for over- and underdense structures.

4.3 Effect of model choice

A comparison of goodness-of-fit between the Gaussian and the weakly non-linear models for the genus curve does not provide conclusive evidence that the weakly non-linear curve is superior. In most cases, the difference between the models is negligible, though when it is not, the weakly non-linear curve is much to be preferred. However, the strongest potential confirmation of the weakly non-linear theory – that the quality of the fit of this model improves relative to the Gaussian one on progressively smaller scales – cannot be demonstrated based on these results. This reinforces the claim that the effect on the genus curve of accounting for weak departures from Gaussianity is modest, though one must expect this will not hold as such departures grow in amplitude.

The results could be taken to indicate that this effect will be best seen on scales smaller than those that have been probed and requires somewhat larger volumes than are available on these scales with the present data. An interesting alternative is to question the impact of

differing forms of galaxy bias on the genus curve. The 2dFGRS has been used (with statistics other than the genus curve) to study the linear and quadratic bias of galaxies (Verde et al. 2002) as well as the stochasticity of the relative bias between early- and late-type galaxies (Wild et al. 2005); topological statistics may be sensitive to a third characterization of bias as *non-local*, as the smoothed distribution incorporates information from regions around the sites of individual galaxies. These will certainly lead to asymmetries in the genus curve of the kind observed here, though it is not clear whether these will involve different modifications to those incorporated in the formalism of equation (15).

It is also possible to investigate the effect of using the weakly non-linear genus curve model when measuring the parameters $\Delta\nu$, A_c and A_v. The dashed lines in Fig. 7 show the value of these parameters when measured relative to the weakly non-linear genus curve with skewness parameters selected as a function of scale. No distinction is to be made between the models in the SGP slice, while the values of the parameters in the NGP slice differ at those scales where the weakly non-linear model provides a substantially better fit.

4.4 Comparison with larger scales

James et al. (2007) have studied the topology of large-scale structure in the 2dFGRS on scales ranging from 8 to 14 h^{-1} Mpc, using slightly different methodology to that employed here with regard to the construction of statistically complete samples, as well as in the parametrization of asymmetries in the genus curve. Fig. 8 extends the results in this work to larger scales by calculating $\Delta\nu$, A_c and A_v for the earlier results.

The two sets of results are consistent at the overlapping scale of 8 h^{-1} Mpc, particularly considering the methodological differences. The cluster parameter in the SGP is a possible exception to this, and indeed it seems clear that little inference about the scale dependence of this statistic across the full range is possible given the data
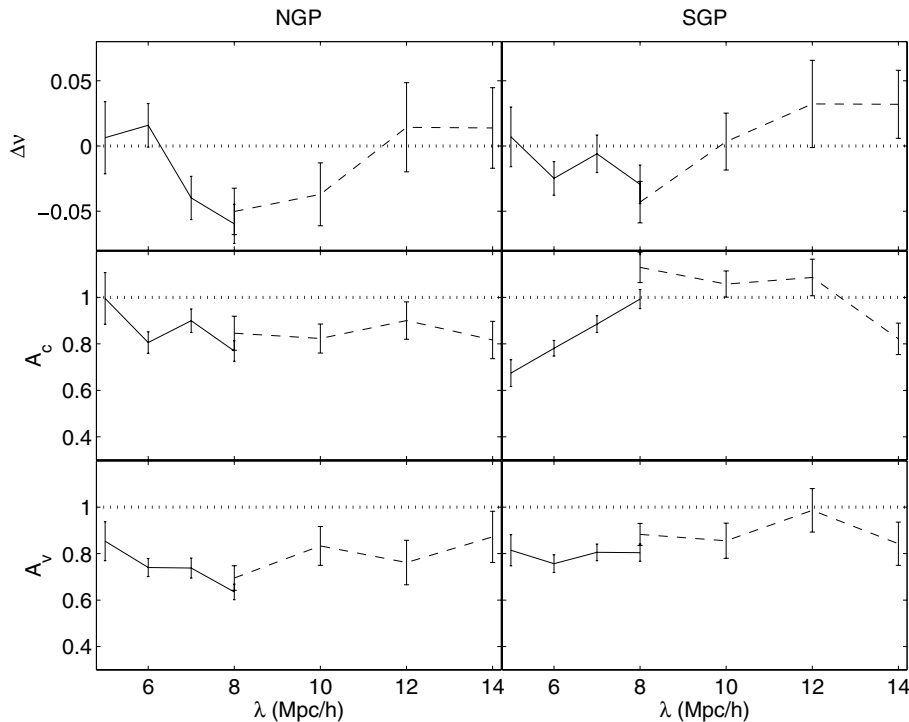

Figure 8. Comparison of metastatistics $\Delta\nu$, A_c and A_v for the 2dFGRS NGP and SGP survey slices between the scales studied in this paper (solid lines; identical to Fig. 7) and those of James et al. (2007) (dashed lines).

presented here. The trends that do emerge are (i) the shift parameter is consistent with zero at larger scales, drops below zero at around $10\,h^{-1}$ Mpc and returns to the smallest scales studied in this work and (ii) the void parameter changes remarkably little when moving into the non-linear regime and is consistent with $A_v \approx 0.8$ across all scales and between both regions of the survey. This value for the void parameter is also seen in the topology of the Sloan subregions reported by Gott et al. (2008), as well as their measurements on the Millennium Run (Springel et al. 2005) and on simulations of dark matter haloes populated with galaxies (Park et al. 2005; Kim & Park 2006), to the point where it seems to deserve recognition as a generic feature of genus measurements. Further work in this area should aim to characterize the scale at which the shift parameter departs from zero and, if the turn-around at lower scales is reproduced in independent samples, the scale at which the meatball shift is maximized. Though the void parameter is now routinely observed to depart from unity, no complete explanation for this observation has emerged. The understanding of both of these trends would profit from examination of the differences between particular galaxy populations. The continued progress of the Sloan Digital Sky Survey and the early results reported by Park et al. (2005) auger well for this effort. Forthcoming surveys based on photometric redshifts are likely to make a substantial contribution through the two-dimensional genus statistic in redshift slices.

It is interesting to speculate on the correspondence between the angular- and three-dimensional correlation functions and the two- and three-dimensional genus statistics, as well as between the ratioing process of Melott & Dominik (1993) and the common practice of random point catalogues to account for geometric and astronomical selection effects in the calculation of correlation functions. Progress on the interpretation of the genus measurements may also require the use of alternative metastatistics. A possible way forward is to use the orthogonality relations of the Hermite polynomials and decompose the genus curve into a series of Hermite functions $\psi_n(\nu) \propto \exp(-\nu^2/2) H_n(\nu)$. The coëfficients of the decomposition are linked to the skewness parameters of the field – in the linear regime only the $n = 2$ coëfficient is non-zero, while in the non-linear regime the $n = 1$ and 3 functions contribute with some small amplitude to the genus curve, inducing a slight asymmetry. A full decomposition of the curve on scales at and below those studied in this work would guide theoretical interpretation and prediction of the topology of non-linear structure.

5 CONCLUSIONS

We have devised a method for measuring the genus curve on flux-limited samples from galaxy redshift surveys in the non-linear regime, setting out the requirements for avoiding major systematic effects in the form of a joint constraint diagram. This method has been applied to the 2dFGRS, providing a smoothed galaxy distribution on which the topology of large-scale structure in the weakly non-linear regime, from 5 to $8\,h^{-1}$ Mpc, has been studied. These measurements are the shortest-scale study of structure topology carried out to date with observational data and have been aimed explicitly at detecting the effects of non-linear structure formation on the genus statistic.

By measuring the genus shift, void multiplicity and cluster multiplicity statistics of the genus curve, we have aimed to quantify departures from the linear regime case of a Gaussian random field. We have observed the effects of structure formation on the genus curve in the form of non-linear gravitational evolution, as demonstrated by a trend in genus shift, and demonstrated that both the cluster and

void multiplicity statistics tend to fall below unity across a broad range of scales. The independence with scale of the latter quantity in particular suggests effects external to the non-linear evolution of the dark matter field, in agreement with earlier suggestions about a relationship between A_v and galaxy bias.

We have studied the effect of model choice in the measurement of these statistics, using analytic predictions for both the linear regime and weakly non-linear regime as comparators to the data. We find that neither model adequately captures the shape of the genus curve, though there is evidence at some scales that the weakly non-linear model provides a substantially better fit. Together, these measurements demonstrate the efficacy of studying topology in the non-linear regime and, given the emphasis future surveys are placing on higher order statistics, suggest that topology will play an important role in answering questions about the formation and evolution of galaxy clusters and dark matter haloes on these scales.

REFERENCES

Adam C., Tannery P., 1996, Œuvres des Descartes. Librairie Philosophique J. Vrin
Adler R. J., 1981, The Geometry of Random Fields. The Geometry of Random Fields. Wiley, Chichester
Albrecht A., Steinhardt P. J., 1982, Phys. Rev. Lett., 48, 1220
Atiyah M. F., Singer I. M., 1963, Bull. Am. Math. Soc., 69, 422
Bardeen J. M., Steinhardt P. J., Turner M. S., 1983, Phys. Rev. D, 28, 679
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Barrow J. D., Bhavsar S. P., Sonoda D. H., 1984, MNRAS, 210, 19
Bennett C. L. et al., 1994, ApJ, 436, 423
Blair M., Gilmore G., 1982, PASP, 94, 742
Canavezes A. et al., 1998, MNRAS, 297, 777
Cole S. et al., 2005, MNRAS, 362, 505
Colless M. et al., 2001, MNRAS, 328, 1039
Croton D. J. et al., 2004a, MNRAS, 352, 1232
Croton D. J. et al., 2004b, MNRAS, 352, 828
Doroshkevich A. G., 1970, Astrophys., 6, 320
Efron B., 1982, CBMS–NSF Regional Conf. Ser. Appl. Math. The Jackknife, the BootStrap and Other Resampling Plans. Society for Industrial and Applied Mathematics (SIAM), Philadelphia
Frigo M., Johnson S. G., 2005, Proc. IEEE, 93, 216
Geller M. J., Huchra J. P., 1989, Sci, 246, 897
Gott J. R., III, Melott A. L., Dickinson M., 1986, ApJ, 306, 341
Gott J. R., III, et al., 1989, ApJ, 340, 625
Gott J. R., III, Mao S., Park C., Lahav O., 1992, ApJ, 385, 26
Gott J. R., III, et al., 2008, ApJ, 675, 16
Guth A. H., 1981, Phys. Rev. D, 23, 347
Hamilton A. J. S., 1988, PASP, 100, 1343
Hamilton A. J. S., Gott J. R., III, Weinberg D., 1986, ApJ, 309, 1
Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399

Hawkins E. et al., 2003, MNRAS, 346, 78

Hikage C. et al., 2002, PASJ, 54, 707

Hikage C. et al., 2003, PASJ, 55, 911

Hikage C., Coles P., Grossi M., Moscardini L., Dolag K., Branchini E., Matarrese S., 2008, MNRAS, 385, 1613

Hogg D. W., 1999, preprint (astro-ph/9905116)

Hoyle F. et al., 2002, ApJ, 580, 663

Hoyle F., Vogeley M. S., Gott J. R., III, 2002, ApJ, 570, 44

James J. B., 2005, Honours Dissertation, Univ. Sydney

James J. B., Lewis G. F., Colless M., 2007, MNRAS, 375, 128

Juszkiewicz R., Weinberg D. H., Amsterdamski P., Chodorowski M., Bouchet F., 1995, ApJ, 442, 39

Kerscher M., Schmalzing J., Buchert T., 1996, in Coles P., Martinez V. J., Pons-Borderia M.-J., eds, ASP Conf. Ser. Vol. 94, Mapping, Measuring and Modeling the Universe. Astron. Soc. Pac., San Francisco, p. 247

Kim J., Park C., 2006, ApJ, 639, 600

Kofman L., Bertschinger E., Gelb J. M., Nusser A., Dekel A., 1994, ApJ, 420, 44

Liddle A. R., Lyth D. H., eds, 2000, Cosmological Inflation and Large-scale Structure. Cambridge Univ. Press, New York

Linde A., 2005, New Astron. Rev., 49, 35

Martínez V. J., Saar E., 2002, Statistics of the Galaxy Distribution. Chapman Hall/CRC, Boca Raton, FL

Martínez V. J., Starck J.-L., Saar E., Donoho D. L., Reynolds S. C., de la Cruz P., Paredes S., 2005, ApJ, 634, 744

Matsubara T., 1994, ApJ, 434, L43

Matsubara T., 2003, ApJ, 584, 1

Matsubara T., Suto Y., 1996, ApJ, 460, 51

Matsubara T., Yokoyama J., 1996, ApJ, 463, 409

Mecke K. R., Buchert T., Wagner H., 1994, A&A, 288, 697

Melott A. L., Dominik K. G., 1993, ApJS, 86, 1

Melott A. L., Weinberg D. H., Gott J. R., III, 1988, ApJ, 328, 50

Melott A. L., Cohen A. P., Hamilton A. J. S., Gott J. R., III., Weinberg D. H., 1989, ApJ, 345, 618

Norberg P. et al., 2002, MNRAS, 336, 907

Park C., Gott J. R., III, 1991, ApJ, 378, 457

Park C., Gott J. R., III, da Costa L. N., 1992, ApJ, 392, L51

Park C., Gott J. R., III, Choi Y. J., 2001, ApJ, 553, 33

Park C., Kim J., Gott J. R., III, 2005, ApJ, 633, 1

Park C. et al., 2005, ApJ, 633, 11

Peacock J. A., 1999, Cosmological Physics. Cambridge Univ. Press, Cambridge

Peebles P. J. E., 1993, Principles of Physical Cosmology. Princeton Series in Physics, Princeton University Press, Princeton, NJ

Press W. H., Schechter P., 1974, ApJ, 187, 425

Sasaki C., 2003, Descartes's Mathematical Thought. Boston Studies in the History and Philosophy of Science, Vol. 237. Springer, Berlin

Schmalzing J., Buchert T., 1997, ApJ, 482, L1

Seto N., Yokoyama J., Matsubara T., Siino M., 1997, ApJS, 110, 177

Shectman S. A., Landy S. D., Oemler A., Tucker D. L., Lin H., Kirshner R. P., Schechter P. L., 1996, ApJ, 470, 172

Spergel D. N. et al., 2006, ApJS, 170, 377

Springel V. et al., 2005, Nat, 435, 629

Tytler D. et al., 2004, ApJ, 617, 1

Verde L. et al., 2002, MNRAS, 335, 432

Vogeley M. S., Park C., Geller M. J., Huchra J. P., Gott J. R., III, 1994, ApJ, 420, 525

Weinberg D. H., Gott J. R., III, Melott A. L., 1987, ApJ, 321, 2

Wild V. et al., 2005, MNRAS, 356, 247

York D. G. et al., 2000, AJ, 120, 1579

This paper has been typeset from a TEX/LATEX file prepared by the author.